REPORT DOCUMENTATION PAGE

AD-A254 138

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 30APR92	3. REPORT TYPE AND DATES COVERED Final 1SEPT90 to 28FEB91

4. TITLE AND SUBTITLE

As on cover sheet

5. FUNDING NUMBERS

F49620-90-C-0077

65502F
3005
A1

6. AUTHOR(S)

Gevins, A. S.
Leong, H. M. F.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

SAM Technology, Inc.
51 Federal Street, Suite 402
San Francisco, CA 94107

8. PERFORMING ORGANIZATION REPORT NUMBER

AEOSR-TR· 92 0009

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Directorate of Life Sciences
AFOSR/NL Building 410
Bolling AFB, DC 20332-6448

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

DTIC
ELECTE
AUG 19 1992
S D

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

AFOSR   Approved for public release;
distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Limitations in people's ability to process and respond to information have become a limiting factor in advanced military aircraft systems. Accordingly, the USAF OSR has been sponsoring research on measuring mental workload as a prerequisite to developing cockpit systems which take the pilot's mental state into account in optimizing overall system performance. During Phase I, we performed a feasibility study in which we analyzed physiological data from four USAF fighter test pilots in search of ways to distinguish between two laboratory tasks which had the same stimulus and response components but differed in level of mental workload. Several electrophysiological measures, alone and in combination, were investigated for their discriminating power including regional brain electrical activity, scalp muscle potentials, and heart and eye activity. Measures were restricted to those which could be recorded in the cockpit, and, in the case of brain signals, to those least likely to be contaminated by head, body and eye movement artifacts. Using a neural network algorithm, we achieved an average of 97% accuracy in classifying independent testing data for the four subjects as either high or low mental workload. Although the results should be interpreted cautiously because of the small number of subjects and the use of artificial laboratory tasks, they suggest that further research will make useful progress in developing a physiological metric of mental workload suitable for use in the cockpit.

14. SUBJECT TERMS

Mental Workload Electrophysiology

15. NUMBER OF PAGES

49

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

Although the results should be interpreted cautiously because of the small number of subjects and the use of artificial laboratory tasks, they suggest that further research will make useful progress in developing a physiological metric of mental workload suitable for use in the cockpit.

92-23073

92 8 18 094

MENTAL WORKLOAD ASSESSMENT IN THE COCKPIT:
FEASIBILITY OF USING ELECTROPHYSIOLOGICAL MEASUREMENTS

PHASE I FINAL TECHNICAL REPORT

AFOSR Contract F49620-90-C-0077
1 SEPT90 to 28 FEB91

PREPARED FOR

Dr. John Tangney
Directorate of Life Sciences
Air Force Office of Scientific Research
Bolling AFB, D.C. 20332-6448

APPROVED BY

Alan Gevins, President

Table of Contents

MENTAL WORKLOAD ASSESSMENT IN THE COCKPIT: FEASIBILITY OF USING ELECTROPHYSIOLOGICAL MEASUREMENTS

PHASE I FINAL REPORT

I. PERSONNEL

Project Period: 90SEPT--91FEB
Alan Gevins, Principal Investigator, 58 hours
Mark Filidei, Research Associate, 477 hours
Tom Laidig, Programmer, 406 hours
Harrison Leong, Signal Processing Engineer, 220 hours
James Johnston, Biophysicist, 214 hours

(Hours are direct costs to the project. Phase II proposal preparation is an indirect cost.)

II. IDENTIFICATION AND SIGNIFICANCE OF THE PROBLEM

Limitations in people's ability to process and respond to information have become a limiting factor in advanced military aircraft systems. Accordingly, the USAF OSR has been sponsoring research on measuring mental workload as a prerequisite to developing cockpit systems which take the pilot's mental status into account in order to optimize overall system performance (Gomer et al., 1979; Wickens, 1979; O'Donnell and Eggemeier, 1988).

The sources and types of information demanding a pilot's attention have increased greatly over the last 30 years (Sexton, 1988). Managing this information intelligently has become crucial in preventing degradation of pilot performance due to excessive mental workload. At the other end of the spectrum, low mental workload during long periods of routine or automated flight is also a problem (Nagel, 1988; Wiener, 1988). This suggests that man-machine system performance could be improved by monitoring a pilot's mental workload and increasing or decreasing task demands to maintain optimum workload levels. Advances in cockpit automation and information display technology have provided ways to accomplish this, but these capabilities cannot be fully exploited for lack of a suitable measure of mental workload. Hence, research on mental workload measurement has become an important topic for human factors engineers, psychologists, and lately, cognitive neuroscientists.

Both theoretical and practical considerations have made it difficult to devise mental workload measures suitable for use in the cockpit. A task's mental load depends on properties of the task, the mental strategy used to perform it, and the capacities of the neural processes underlying perception, thought and decision, and motor control. The mental resources required to execute a task can change in importance and kind as mental strategies change with experience (Natani and Gomer, 1981). These resources can have have nonlinear interactions as well (Gevins, 1989b; Freeman and Skarda, 1985; Freeman, 1983). Hence, characterizing how mental resources are used and the relationship to overall mental load is, even conceptually, a difficult problem. Practical measures for assessing mental load in the cockpit have the added problem of being sensitive to irrelevant factors. For example, measurements based on the electroencephalogram (EEG) are sensitive to head, jaw, limb and other movements that may or may not be relevant to the mental load of the task at hand.

In the laboratory, much progress has been made in quantifying mental workload (for reviews, see Boff et al., 1986; Boff and Lincoln, 1988). Many different methods of measuring mental workload have been explored (see Appendix B), including: subjective estimates; direct task performance measures; secondary task performance; speech characteristics; and physiological measures such as pupil response, evoked potentials (primarily P300), electrocardiograms (EKGs), electroencephalograms, skin conductance, eye movements and blinks, and respiration. Most of the methods are not suitable for the active cockpit and other field situations in which near real-time assessment is needed and high load episodes

are frequent and unpredictable. The best candidates seem to be electrophysiological measurements, specifically continuous EEG, eye blinks and movements, scalp muscle potentials, heart rate, and respiration. Unlike other candidate measures, these are continuously available, do not restrict or alter task structure, and do not introduce additional workload. The measures are complementary; some are relatively specific and some are non-specific with regard to particular mental resources (Andreassi, 1989). Thus use of several types of measure may provide an ability to index mental workload in a wide variety of situations. In addition, technology for measuring these signals can be unobtrusively integrated into the active cockpit environment as an integral part of a flight suit and helmet (Albery and Van Patten, 1991; Lewis et al., 1988; Cammarota, 1990, 1991).

Though perennially promising, a review of the psychophysiological workload literature reveals that much research is needed before practical metrics can be developed. In most studies reviewed in Appendix B, conceptual problems or practical considerations limit the generalizability of indices found. For example, in the case of EEGs, it is particularly easy to be deluded into thinking that one is measuring varying levels of mental workload, when one in fact is measuring cortical signals associated with varying amounts of limb movements. Since it is well known that frontal and central EEG low frequency power increases with increased motor response activity (Gevins and Schaffer, 1980), power in these bands can appear to index mental workload in tasks in situations in which mental workload levels also differ in amount of hand movements, eg. easy and hard levels of many video games. Such an index would not generalize to tasks where increased mental workload was not associated with increased limb movements and, at worst, could be fooled by movements that are unrelated to task performance. In addition, an index that depended on EEG low frequency activity would not work in practice since head and eye movement artifacts are strong in this band; in an actual or simulated flight situation, these movements are prodigious. For these reasons, in our Phase I feasibility study, low frequency EEG activity from frontal and fronto-central scalp locations were excluded from consideration.

Although laboratory studies of workload are important prerequisites to developing a mental workload index, we believe that extension to more realistic tasks is a necessary step which has not received adequate attention. Our intent is to remedy this with a three-pronged approach consisting of developing improved data acquisition technology, developing effective automatic artifact processing techniques, and developing methods to construct and apply mental workload indices based on continuous electrophysiological signals. The research discussed here is concerned with the third topic.

Current methods to obtain electrophysiological recordings require substantial preparation of the subject and, hence, are impractical for routine use. With funding from USAFSAM, we are developing an electrophysiological recording system built into a flight helmet that requires no preparation of the scalp. The system is currently being tested and refined.

Artifacts from head, body, limb, eye, and other motor activities often hide useful components of electrophysiological signals. For example, the spectral characteristics of these artifacts can overlap those of cognitive-related EEG components. Consequently, considering the level of physical activity in the cockpit, especially during periods in which mental workload assessment would be most useful, developing automatic techniques to effectively detect and, when possible, correct artifacts is crucial to constructing a practical system.

Methods to construct workload indices constitute the third component of our approach. The methods would need to handle the signals output by our in-flight recording system after they have been classified as clean by the automatic artifact processing system. Here we describe the initial feasibility test to discriminate two mental workload levels using neural network pattern recognition methods. The significance of the work is that most prior workload studies have measured a single variable derived from a single physiological measurement, and have applied standard linear statistical tests (typically Analysis of Variance) to test for significant differences across mental workload levels. Our results suggest that linear statistical methods are suboptimal for this problem. We observed that combining several types of physiological measurements, possibly using several variables derived from each,

substantially enhanced the ability to discriminate mental workload levels. The multivariate approach is in accordance with the view that mental workload is differentially expressed by many physiological subsystems spanning both the central and autonomic nervous systems. Another important benefit of using a combination of measures is reliability: Signals from one modality may be temporarily unusable due to artifact while another modality remains clean; e.g., jaw clenching may contaminate EEGs, but eye movement measures would be unaffected.

III. OBJECTIVE

Identify regional EEG features, scalp muscle activity features, eye blink features, and heart rate and heart rate variability features that, in combination, are best at distinguishing between the performance of two workload levels of a laboratory visuomotor memory task. Constrain the analysis to signal features and signal processing methods that would potentially be practical to use in real-time assessment of mental workload of aircraft fighter pilots in action.

IV. STATUS OF RESEARCH EFFORT

A previous experiment on sustained mental work (Gevins et al., 1988a, 1990; see Appendix C) gave us the opportunity to analyze EEG, EKG, and eye blink data that were recorded while USAF fighter test pilots performed two laboratory tasks that differed only in the amount of mental effort required.

1. *Description of the Experiment*

We analyzed data recorded from four right-handed, male USAF fighter test pilots. The pilots practiced a brief battery of tasks, including a visuomotor memory task at two difficulty levels, for about six hours, until the learning curves for response time and error stabilized. Subjects began at about 13:30 the following day, and performed the task battery during the ensuing 10 to 14 hours. The session consisted of a 5-8 hour work period, a brief dinner break, then another 5-7 hour work period which ended when the subject was too exhausted to continue. We analyzed visuomotor-memory task data collected between 13:30 and 20:30, before subjects showed subjective, behavioral or neural signs of fatigue.

1.1. *The Visuomotor-Memory Task*

The less difficult level of mental workload, called the zero-back task, required pilots to respond to a visually presented numeric digit with a precise finger pressure on an isometric pressure transducer in proportion to the numeric value of the stimulus. No response was to be made when the value of the stimulus was zero, which occurred in roughly 20% of the trials, randomly distributed. The more difficult level of mental workload, called the two-back task, required a response with a finger pressure proportional to a visual stimulus that had appeared two trials back. No response was to be made when the current stimulus number was the same as the two-back number, which occurred in roughly 20% of the trials, randomly distributed. For example, if the stimulus sequence were 9, 7, 6, 3, 6 they would need to repond with finger pressure .9 kg to the first 6, .7 kg to the 3, and would not respond to the second 6. Each trial consisted of a cue, which consisted of the disappearance of an X centered on the video screen, a 100 msec numeric stimulus which appeared 750 msec following the cue, a response, and one second after the response, feedback consisting of a two digit number characterizing the accuracy of the response. Pilots were instructed to try to only blink during the inter-trial interval when a dot was on the screen.

1.1.1. *Relationship to Mental Workload*

Note that these two tasks had exactly the same stimulus characteristics and required exactly the same type of responses, namely an isometric finger pressure response with minimal overt movement. The tasks thus differed only in the level of mental workload. In the zero-back condition, pilots simply needed to produce a graded pressure after evaluating the stimulus. The two-back condition was more complex: pilots had to remember the two previous numbers in the presence of numeric distractors (the feedback stimuli), evaluate the current stimulus to determine if a response was actually required, and produce the graded pressure response when required.

1.2. *Physiological Measurements*

The following signals were recorded: EEG from 27 electrodes referenced to the right mastoid, vertical (VEOG) and horizontal (HEOG) eye movements, EMG activity of the right flexor digitorum muscles, EKG, respiration, and EEG activity at the left mastoid. Signals were digitized beginning with the get-ready cue and continued through 1.5 seconds following feedback stimuli. All signals were amplified by a Bioelectric Systems Model AS-64P amplifier with 0.016 to 50 Hz passband and digitized to 11 bits at 128 Hz. The reference for EEG signals was converted to digitally linked mastoids.

2. *Analysis*

The main goal of the analysis was to find features based on electrophysiological signals that would accurately distinguish 0-back trials from 2-back trials. Two subgoals guided our choices in this analysis: 1) to simulate signal analysis appropriate to workload measures which could be made in-flight, and 2) to determine the usefulness of multimodality signals to quantify mental workload.

We achieved the first subgoal by training and testing pattern classifiers with overlapped sets of trials. This simulated using a sliding window of data to obtain a continuous estimate of workload where, within this window, portions of the signal would not be used because of contamination. In addition, we did not consider EEG signal features below 4 Hz since, in the cockpit, highpass filters with approximately a 4 Hz cut-off would be required to reduce or eliminate ubiquitous head and body movement artifacts. We achieved the second subgoal by investigating the classification power of EEG, EMG, and EKG based features alone and in combination. Eye movement features were not used because there was insufficient data, resulting from the instruction to subjects to blink only during inter-trial intervals.

2.1. *Overview*

The Phase I analysis consisted of removing trials with contaminated data, making separate training and testing data sets, choosing signal features using prior knowledge about the problem, computing feature values on the training data, examining the distributions of these values to choose a set of candidate features for classifier-directed feature selection and classification analysis, performing these analyses for several candidate feature sets, and validating classification performance using an independent subset of data. In accord with our belief that a viable workload measure will need to be adjusted to each person, the analyses were done independently for each subject.

2.2. *Artifact Processing*

The EEG data were reviewed and edited for gross head and body movement artifacts, excessive muscle contamination, eye movement and blinks, artifacts due to poor electrode contact, and dead or saturated channels. EKG signals were contaminated with EMG bursts and varying degrees of movement. EKG

data was reviewed and marked for contaminants during the extraction of interbeat intervals (IBIs), the interval between successive R-waves. Trials with less than 2 IBIs or with incorrect R-wave detections were deleted.

2.3. Signal Features

2.3.1. EEG

Based on past work (Gevins, 1979abc, and Appendix B), we searched for spectral bands for which power differed between the two workload levels. To do this, power spectra were computed for each trial windowed with a 25% cosine taper. Each trial had roughly 3-4 seconds of data. We examined average spectra for 16 channels of EEG (see Figures 1 through 4 in Appendix A).

After reviewing these plots, we choose to use fairly standard EEG bands: theta (4 to 7 Hz), alpha (8 to 13 Hz) and beta1 (14 to 25 Hz). Our previous work suggested that both alpha and beta band power decrease with increasing workload, while theta band power increases (Gevins et al, 1979abc). Thus, not finding anything in the spectral plots to indicate otherwise, we chose these standard spectral power bands for the EEG feature domain. Hanning-windowed FIR filters with 0.5 second impulse response length (6 dB down at the edge frequencies; 17.8 db/octave rolloff) were constructed and applied to each trial.

2.3.2. EKG

Two features were extracted from the EKG data: heart rate (HR) based on IBIs and heart rate variation (HRV) based on the root mean square of successive differences between successive IBIs (Heslegrave, et al., 1979). Because no data were collected between trials, we could only estimate HR and HRV within trials; typically, estimates were based on 3 to 4 IBIs. To measure IBI, our program detected R-waves by using a weighted average with a 3-5 second time constant for an adaptive baseline offset, and an adaptive threshold based on a slower average of already detected R-peaks. Timing constraints were used to ignore peaks too close together, or to ignore IBIs that were too long. Moderate baseline variation within a trial and moderate EMG bursts were handled well by the detector.

2.3.3. EMG

Scalp EMG features were generated from a 25-55 Hz, Hanning-windowed FIR band-pass filter with 0.5 second impulse length, applied to lateral and frontal peripheral EEG channels.

2.3.4. EOG

We examined waveform features of vertical eye signals (VEO) following recent work on eye blinks and workload (Morris, 1984ab, 1985; Skelly, et al., 1987; Wilson, et al., 1987). We could not examine eye blink rate measures (e.g., Stern and Skelly, 1984) since our data were not continuously recorded. Features included peak amplitude, total blink duration (defined as peak width at 50% peak amplitude), area (computed as peak amplitude times duration), aspect ratio (the ratio of peak amplitude over total duration), and asymmetry (trailing edge duration over leading edge duration). To compute these features, VEO data were filtered through a 15-Hz, Hamming-windowed, FIR lowpass filter with an impulse response length of 0.0625 sec. After filtering, the feature detector located the peak signal level within each trial, and labeled that point as a blink candidate. All timepoints where the signal magnitude was less than 5% of the peak level were considered to be part of the baseline, and a linear least-squares fit was performed on these points to estimate baseline offset and drift (trend). The feature

detector subtracted this estimate, recalculated which points met the 5% criterion, and iterated until no significant change occurred. The detector then computed the width of the blink by finding the 50% points on the leading and trailing edges of the blink. With the peak point, these points defined the total duration of the blink, and the duration of the leading and trailing edges.

Following these computations, we found that there were an insufficient number of events for classification analysis; we therefore dropped these features from further consideration.

2.3.5. Windowed Features

All raw feature values were converted to windowed features by computing means and variances over successive windows of n consecutive trials. Windows overlapped by $n-1$ trials. This high degree of overlap corresponds to an updated workload measurement roughly every 3-4 seconds with a time resolution of roughly $3n$ to $4n$ seconds.

We were careful to separate trials into training and test sets before applying the windowing operation. Hence, classifiers were guaranteed to be tested on data independent of that used to construct them. Classification results could have been biased because data samples were highly correlated due to the high degree of overlap. So, as a further precaution, we tested classifiers with windowed features computed with successive non-overlapped windows; this was done only when n was sufficiently small to provide a reasonable number of testing samples.

2.4. Selecting Features for Neural Network Classification Analysis

Our general approach was to compute a one-dimensional classification error probability for each candidate mean and variance feature and select those features with low error probabilities, taking care to select a group of features that had a good representation of electrode locations and frequency bands and excluding features that might be overly prone to artifact (e.g., frontal theta from tiny eye movements). Error probabilities were estimated by computing z-scores across workload conditions, estimating the distributions of these values, and finding a threshold for which classification based on this threshold would result in minimal errors. Figure 5 shows an example of these values across the EEG channels studied.

We divided the pattern recognition study into two parts differing according to the features used: 1) EEG measures unconfounded by non-cognitive processes (e.g., neural control of movement and muscle activity), the "clean neurocognitive signal" (CNS) study, and 2) EEG measures mixed with scalp EMG, the "mixed measures" (MM) study. For the CNS study, we excluded beta-band features from peripheral channels, which are likely to have EMG contamination; frontal and central theta-band features, which are likely to have significant motor control components; and the EMG band (25-55 Hz). For the MM study, we relaxed these constraints. It turned out that we could not include EKG features in the MM study for lack of sufficient trials. (We were surprised to find that the intersection between trials with non-contaminated EEG and trials for which EKG IBIs could be estimated was so small for each subject.)

For one subject, in the CNS study, Principle Components Analysis (PCA) had to be performed to find features with good classification performance. PCA features were examined exactly as non-transformed features were examined. In performing PCA, it was possible to perform some feature selection: we examined the weights of the original features in the PC's which had low error probabilities. Those that had little influence in this subset of PC's were excluded and PCA was performed on the remaining features.

2.5. Neural Network Classification

We used a pattern classification algorithm that, from a set of candidate variables, automatically generates a two-layered, feed-forward neural network; trains and tests the network; and identifies small subsets of variables that produce the best classification (Gevins, 1980; Gevins and Morgan, 1986, 1988; Viglione, 1970; Joseph, 1961). In brief, the Joseph-Viglione algorithm chooses small, unique combinations of candidate variables with which to construct candidate neural units to use in the first layer, the "input" layer, of the network. Discriminant analysis is used to determine characteristics of the candidate units. Initially, the candidate unit with the best classification performance is selected and connected to the single output unit of the network. The connection weight to the output unit and its threshold are adjusted iteratively to minimize classification error. The algorithm continues to add "input" layer neural units one at a time until a pre-specified limit is reached or an addition fails to significantly improve classification accuracy. At each iteration, the algorithm picks the candidate neural unit that maximally improves classification accuracy. This algorithm was given 10 to 20 candidate variables and found subsets of 2 to 4 variables that gave high classification performance.

3. Results

3.1. EKG

Table 1 (all tables appear in Appendix A) reports univariate classification accuracy, 100(1 - error probability), for data processed by a 20-trial window. Mean HR, variance of HR, and mean of HRV are shown. We included all trials for which IBI could be estimated irrespective of the cleanliness of EEG. EKG data was not recorded for subject two. Mean classification performance across EKG derived features for the other three subjects were 63%, 72%, and 70%. The results suggest that the EKG features were not very sensitive indices of mental workload for our laboratory tasks. EKG features may be more sensitive in the cockpit environment where autonomic arousal varies widely.

3.2. Clean Neurocognitive Signal (CNS) Study

High classification results were achieved with a 20-trial window length which translates to roughly 80 second resolution. Results are shown in Table 2. We report test set performance for the simplest neural network that achieved at least 90% classification accuracy in training. Test accuracies for the four subjects were 97%, 100%, 100%, and 92%. The most important features for classification were different for each subject. Temporal theta and/or alpha activity were highly important features for three of the subjects. Occipital theta, frontal alpha, and central beta activity were each highly important for at least one of the subjects. For subject four, principal components were used as input variables to the neural network algorithm. The relative feature weightings reported for this subject are the effective weightings after accounting for the relative importance of each feature in each principal component used and the relative importance of each principal component in the network classifier.

3.3. Mixed Measures Study

Good classification results were achieved with a 5-trial window which translates to roughly 20 second resolution (Table 3). We report test set performance for the simplest neural network that achieved at least 90% classification accuracy in training. Test accuracies for the four subjects were 95%, 100%, 99%, and 94%. Test accuracies achieved for test set data which had been windowed with non-overlapped, 5-trial windows were 92%, 94%, 100%, and 94%. There was insufficient data to determine whether or not there was a significant difference between the two sets of accuracies but they appear to be comparable. Frontal and occipital EMG activity was highly important for three of the subjects. For

subject two, temporal and frontal alpha activity were still the most important features despite the higher time resolution of this study; however, we note that high classification accuracy could not be achieved at this time resolution if scalp EMG was excluded. We are not aware of any prior publications reporting the sensitivity of scalp EMG to mental workload.

4. Discussion

In this small-scale experiment, we determined the feasibility of distinguishing between two levels of mental workload of a laboratory task using subject-specific measures of ongoing EEG and scalp muscle activity (EMG). Since stimulus and response properties were exactly the same between workload levels, and since we did not use measures which are particularly sensitive to head, body and eye movement artifacts, there is a reasonable inference that the physiological measures actually reflect mental workload.

The good results we achieved in both the clean neurophysiological signal (CNS) and mixed measures (MM) studies support the view that mental workload can be indexed in several ways. Which way would be best depends on many factors including the spectral regions in which clean signal would be available, the task contexts in which a mental workload measure would be desired, and the generalizability and sensitivity of the index within these contexts. Our experimental results illustrate one of the trade-offs that would need to be considered. The CNS study was conducted with the thought that an index which depended only on signals that were likely to have a direct relationship with a spectrum of higher cognitive brain functions would be more likely to generalize across tasks since it would not depend on idiosyncratic perceptual and motor demands of a task. Clearly, the trade-off was that time resolution was fourfold less than that achievable by including scalp muscle potential (EMG) measurements in the index. The best practical solution may be a combination of indices, each optimal in a different situation, possibly a requirement if time resolution is to be improved while maintaining sensitivity and generalizability.

Although scalp EMG was found to be an important indicator of mental workload for all subjects, it was also apparent that the finer details of index structure were highly specific to each subject. This is consistent with the view that mental workload consists of multiple effects over a cross section of neural subsystems, each having a different electrophysiological representation. The exact representation depended on each subject's own functional neuroanatomy. In addition, the mental effort required to perform a task depended on each subject's past experience, abilities, and present model of the task. The conclusion that can be drawn is that it is possible to find common factors of mental workload but the particular pattern of expression of mental workload through these factors can differ considerably among individuals. To minimize the effects of these differences on index sensitivity and reliability, any index of mental workload must be capable of being calibrated to an individual.

Results of our MM study are limited to EEG and EMG because the original experiment whose data we analyzed here was not designed to make multi-modal measurements of mental workload. Discontinuity in EKG and respiration signal recordings between trials strongly compromised our ability to adequately estimate EKG features; pilots were instructed to blink between trials when data was not recorded; and task stimuli were small and centrally displayed to minimize eye movements. Our Phase II research will allow us to investigate this possibility by using continuously recorded electrophysiological signals and tasks requiring a wider variety of mental and physical responses.

The most important extension of these results will be to test the methods on other tasks. The tasks used here tested differential loading on working memory; hence, generalization to other mental resources remains unknown. The repetitive nature of our task protocol may also limit generalizability. Since pilots performed very similar mental tasks during each trial, computing means of the spectral features across trials could improve the signal to noise ratio and bring out components of the signals that

happened to be topographically stable across trials. In actual or simulated flight situations, these benefits would not be realized since, in general, there would be a continuously changing variety of mental processes. Hence, the discriminations we achieved may be based on unrealistically "clean" signals and topographical differences that are peculiar to the tasks we chose. On the other hand, the electrophysiological signals used in the CNS study are considered general measures of arousal and concentration and, thus, would have minimal dependence on the particular mental resources and structural details of a task. Also considering that we excluded the possibility of dependence on idiosyncratic perceptual and motor components of the tasks by design, it seems plausible that the indices we found in the CNS study may be truly valid across tasks.

Regardless, in our Phase II research, generalization will be tested directly by using several distinct families of tasks, each of which demands a different mix of cognitive resources, and an appropriate task presentation protocol to sidestep the pitfalls we have noted. Using families of tasks will allow us to develop methods to index more than two levels of mental workload. Candidate indices will be tested for their ability to interpolate or extrapolate levels of mental workload other than those used to derive them. This must be done in parallel with analyzing data collected in flight and in moving base flight simulators to identify practical constraints that must be considered so that indices are robust against the physical and psychological extremes of the cockpit.

V. REFERENCES

Albery, W.B. & Van Patten, R.E. (1991), Non-Invasive Sensing Systems for Acceleration-Induced Physiologic Changes. *IEEE Engr. in Med. and Biol.* pp. 46-51.

Andreassi J.L., (1989), Psychophysiology: Human Behavior and Physiological Response, Lawrence Erlbaum and Associates.

Boff, K.R., Lincoln, J.E., (Eds.), (1988), *Engineering Data Compendium: Human Perception and Performance, Vol. II*, John Wiley and Sons, New York.

Boff, K.R., Kaufman, L., Thomas, J.P., (Eds.), (1986), *Handbook of Perception and Human Performance Vol. II: Cognitive Processes and Performance*, John Wiley and Sons, New York.

Cammarota, J.P. (1991), Integrated System for Detecting and Managing Acceleration-Induced Loss of Consciousness. *IEEE Engr. in Med. and Biol.* pp. 52-55.

Cammarota, J.P. (1990), Evaluation of Full-Sortie Closed-Loop Simulated Aerial Combat Maneuvering on the Human Centrifuge. *Proc. IEEE National Aerospace & Electronics Conference.* pp. 838-842.

Freeman, W.J., (1983), "The physiological basis of mental images," *Academic Address, Biol. Psychiatry*, v18, pp1107-1125.

Freeman, W.J. and Skarda, C.A., (1985), "Spatial EEG patterns, non-linear dynamics and perception: the neo-Sherringtonian view," *Brain Res. Rev.*, v10, pp147-175.

Gevins, A.S. (1980) Pattern recognition of brain electrical potentials. *IEEE Trans. Patt. Anal. Mach. Intell., PAMI-2* (5), pp. 383-404.

Gevins, A.S. & Morgan, N.H. (1988) Applications of neural-network (NN) signal processing in brain research. *IEEE ASSP Trans., 36,* (7), pp. 1152-1161.

Gevins, A.S., Bressler, S.L., Cutillo, B.A., Illes, J., Miller, J., Stern, J., Jex, H. (1990) Effects of prolonged mental work on functional brain topography. *EEG clin. Neurophysiol.*

Gevins, A.S., Bressler, S.L., Morgan, N.H., Cutillo, B.A., White, R.M., Greer, D. & Illes, J. (1989b) Event-related covariances during a bimanual visuomotor task. Part II. v.ul Electroencephalogr. clin. Neurophys., 74(2) 147-160.

Gevins, A.S., Cutillo, B.A., Fowler-White, R.M., Illes, J. & Bressler, S.L. (1988a) Neurophysiological patterns of operational fatigue: preliminary results. *NATO/AGARD Conference Proceedings, 432,* pp. 22-1 to 22-7.

Gevins, A.S. and Morgan, N.H. (1986) Classifier-directed signal processing in brain research. *IEEE Trans. Biomed. Eng., BME-33* (12), pp. 1054-1068.

Gevins A.S. and Schaffer R.E., (1980) "A critical review of electroencephalographic (EEG) correlates of higher cortical functions", CRC Critical Reviews in Bioeng., v4, pp113-164. IP

Gevins, A.S., Zeitlin, G.M., Doyle, J.C., Yingling, C.D., Schaffer, R.E., Callaway, E. & Yeager, C.L. (1979a) EEG correlates of higher cortical functions. *Science, 203,* pp. 665-668.

Gevins, A.S., Zeitlin, G.M., Yingling, C.D., Doyle, J.C., Dedon, M.F., Schaffer, R.E., Roumasset, J.T. & Yeager, C.L. (1979b) EEG patterns during "cognitive" tasks. I. Methodology and analysis of complex behaviors. *Electroencephalogr. clin. Neurophys., 47,* pp. 693-703.

Gevins, A.S., Zeitlin, G.M., Doyle, J.C., Schaffer, R.E., & Callaway, E. (1979c) EEG patterns during "cognitive" tasks. II. Analysis of controlled tasks. *Electroencephalogr. clin. Neurophysiol., 47,* pp. 704-710.

Gomer F.E., Beideman L.R., and Levine S.H., (1979) "The application of biocybernetic techniques to enhance pilot performance during tactical missions", McDonnell Douglas Astronautics Company, St. Louis Division, Tec. Rpt. MDC-e2046.

Heslegrave Ronald J., John C. Ogilvie, and John Furedy (1979) "Measuring baseline-treatment differences in heart rate variability: Variance versus successive difference mean square and beats per minute versus interbeat intervals", Society for Psychophysiological Research, Vol 16, pp. 151-157.

Heslegrave R.J. and J.J. Furedy (1979) "Sensitivities of HR and T-wave amplitude for detecting cognitive and anticipatory stress", Physiology and Behavior, 22, pp. 17-23.

Joseph, R.D., (1961), *Contributions to Perceptron Theory*, Ph.D. Thesis, Cornell Univ., Ithaca, New York.

Lewis, N.L, McGovern, J.B., Miller, J.C, Eddy, D.R. & Forster, E. M. (1988), EEG indices of G-induced loss of consciousness (G-LOC). NATO/AGARD conference proceedings, 432, pp. 29-1 to 29-12.

Morris T.L. (1984a) " Electrooculographic indices of fatigue-induced decrements in flying related performance", (Technical Report). Brooks Air Force Base, TX: School of Aerospace Medicine.

Morris T.L. (1984b) " Electrooculographic measurement, fatigue, and variability of performance in simulated aircraft flight, Dissertation, College Station, TX: Texas A and M University", (Technical Report). Brooks Air Force Base, TX: School of Aerospace Medicine.

Nagel, D.C., (1988), "Human error in aviation operations," in Wiener, E.L., Nagel, D.C., (Eds.), *Human Factors in Aviation*, pp263-303.

Natani K., Gomer F.E., (1981), "Electrocortical activity and operator workload: a comparison of changes in the electroencephalogram and in event-related potentials", McDonnell Douglas Astronautics Company, St. Louis Division, Tec. Rpt. MDC-e2427.

O'Donnell R.D. and Eggemeier F.T., (1988), "Workload assessment methodology", in Boff K.R., Kaufman L., and Thomas J.P. (Eds.), Handbook of Perception and Human Performance VII: Cognitive Processes and Performance, Chap. 42.

Sexton, G.A., (1988), "Cockpit-crew systems design and integration," in Wiener, E.L., Nagel, D.C., (Eds.), *Human Factors in Aviation*, pp495-526.

Skelly J.J., B. Purvis and G. Wilson (1987) "Fighter pilot performance during airborne and simulator missions: Physiological comparisons", Proceedings of the Advisory Group on Aerospace Research and Development (AGARD) Aerospace Medical Panel Symposium, Electric and Magnetic Activity of the Central Nervous System: Research and Clinical Application in Aerospace Medicine.

Stern J.A. and Skelly J.J., (1984), "The eye blink and workload consideration," Proc. Human Factors Soc., v28, pp942-944.

Viglione, S.S. (1970), "Applications of pattern recognition technology," In: J.M. Mendel & K.S. Fu, *Adaptive Learning and Pattern Recognition Systems*, New York: Academic Press.

Wickens C.D., (1979), "Measures of workload, stress, and secondary tasks", in Moray N. (Ed.), Mental Workload, New York, Plenum Press, pp79-99.

Wiener, E.L., (1988), "Cockpit Automation," in Wiener, E.L., Nagel, D.C., (Eds.), *Human Factors in Aviation*, pp433-461.

Wilson, G.F., Purvis, B., Skelly, B., Fullenkamp, P., and Davis, I., (1987), "Physiological data used to measure pilot workload in actual flight and simulator conditions," In *Proc. 31st Ann. Mtg. Human Factors Soc.*

APPENDIX A, Figures and Tables

Table of Contents

—

Figures 1 to 4: Average power spectra of EEG data for the low (0 back -- thin line) and high (2-back -- thick line) mental workload conditions from each subject for all 16 channels examined. Signals of each trial were windowed with a 25% cosine taper before taking the FFT. Spectra averaged over all trials are shown. EEG channel names are shown above each plot. For all subjects but one, one or more channels exhibit depressed alpha activity (around 10 Hz) and increased beta and scalp EMG activity (above 13 Hz) with increased workload. The exception is subject 4 who shows decreased beta and scalp EMG activity with higher workload. These plots helped determine spectral features used in the pattern recognition analysis.

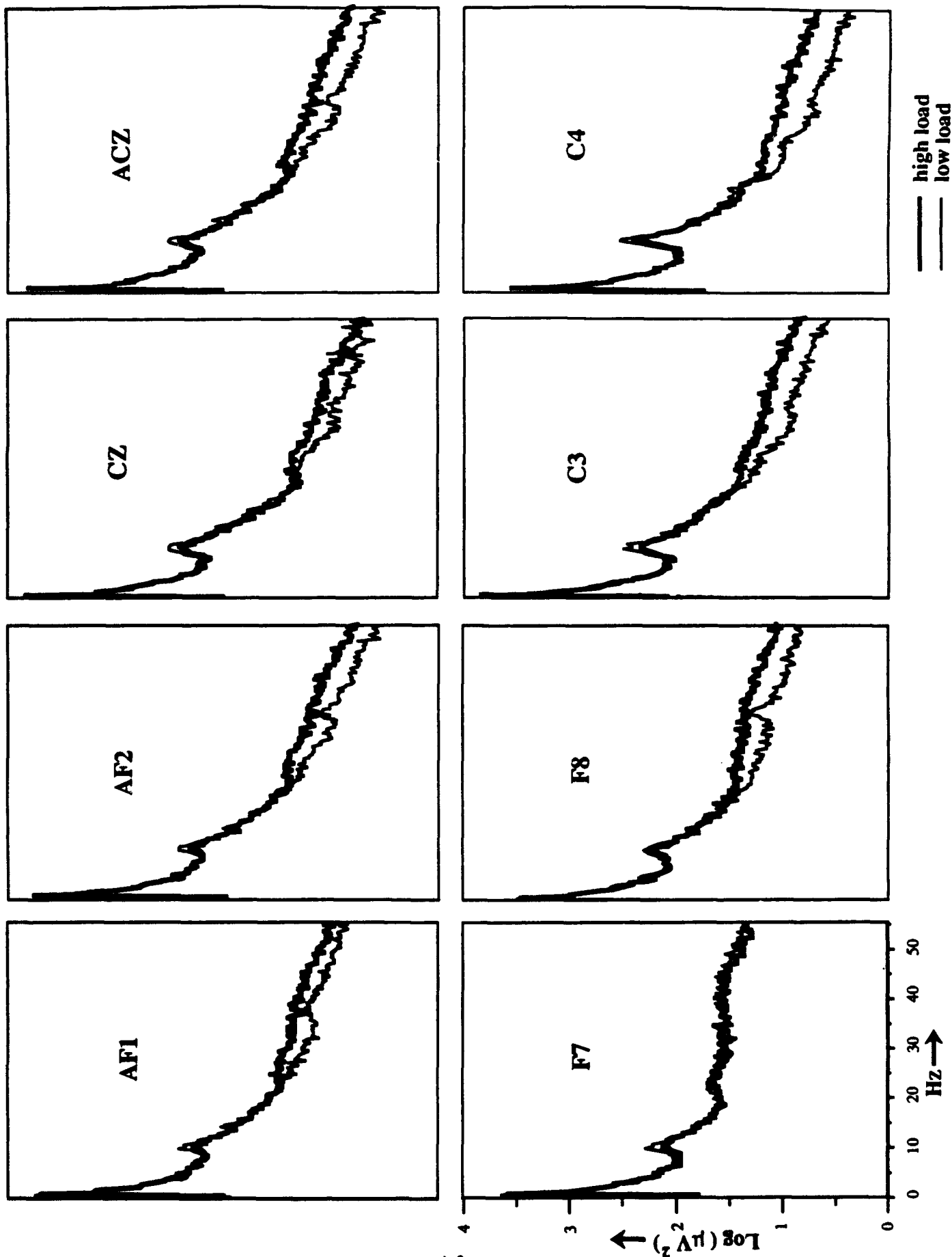**Figure 1a: Subject 1, average power spectra.**

ACZ

C4

CZ

C3

AF2

F8

AF1

F7

high load
low load

Hz →

← Log (μV²)

Figure 1b: Subject 1, average power spectra.

— high load
—— low load
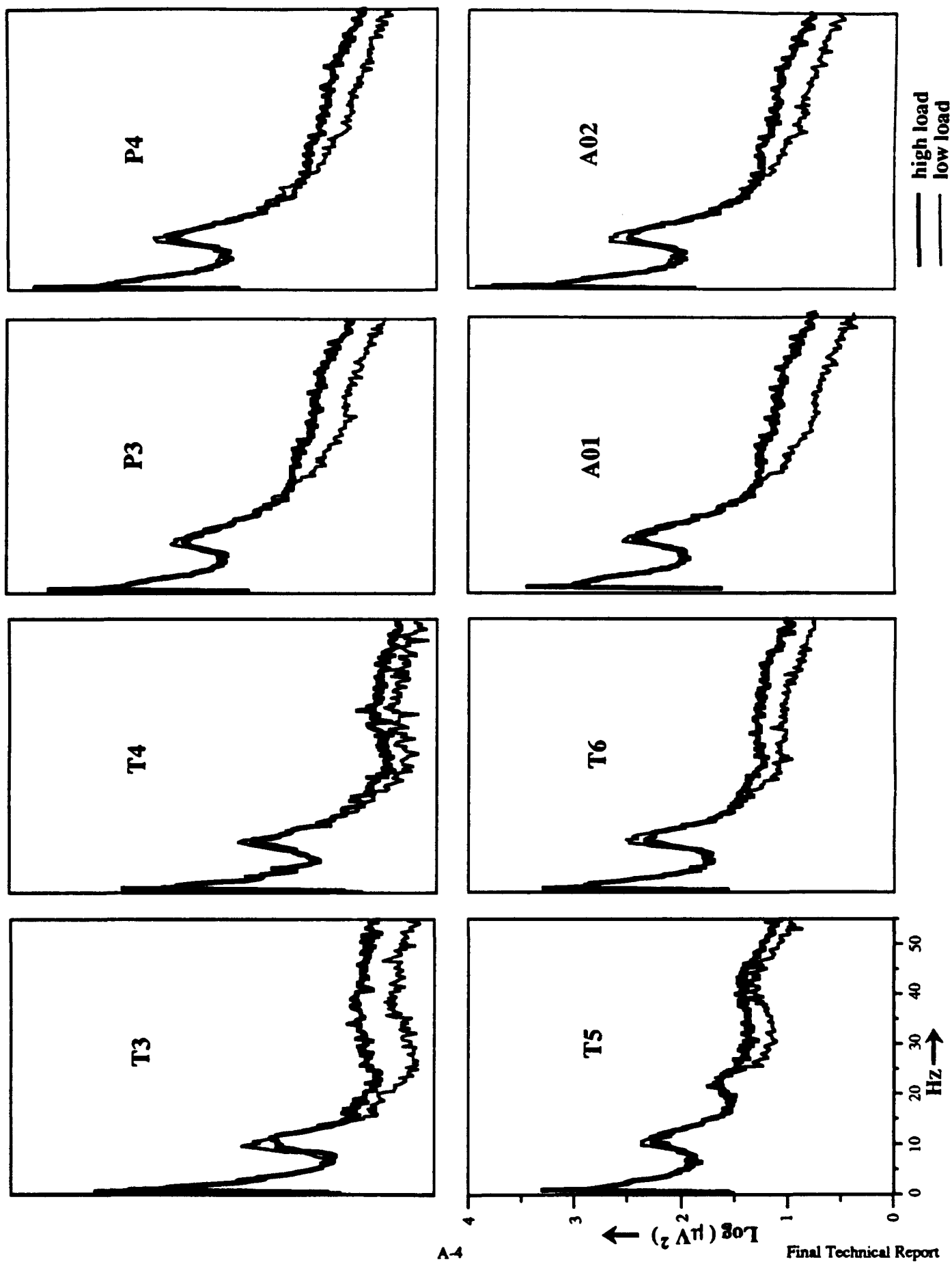
$Log\ (\mu V^2)$

Hz →

Final Technical Report
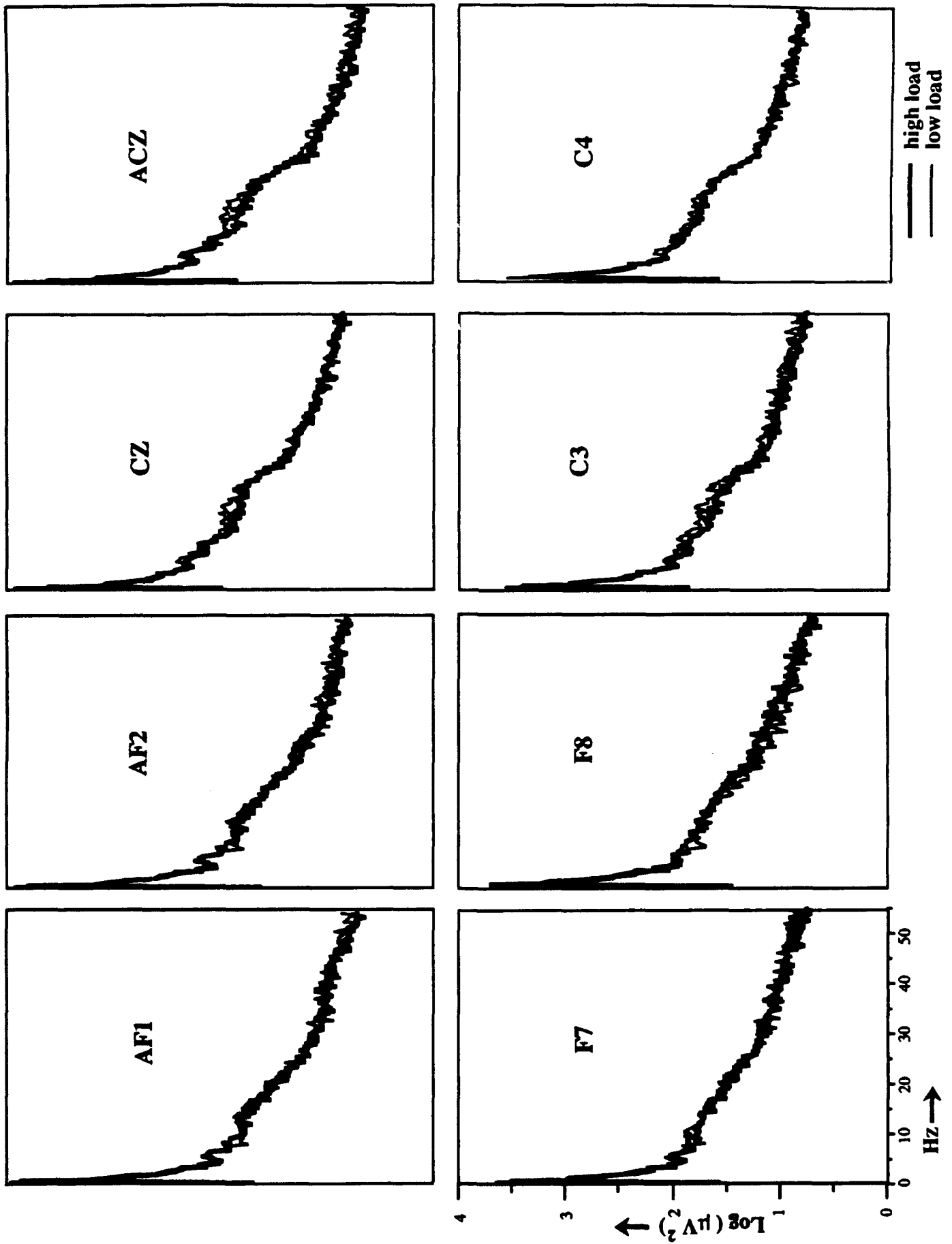
Figure 2a Subject 2, average power spectra.
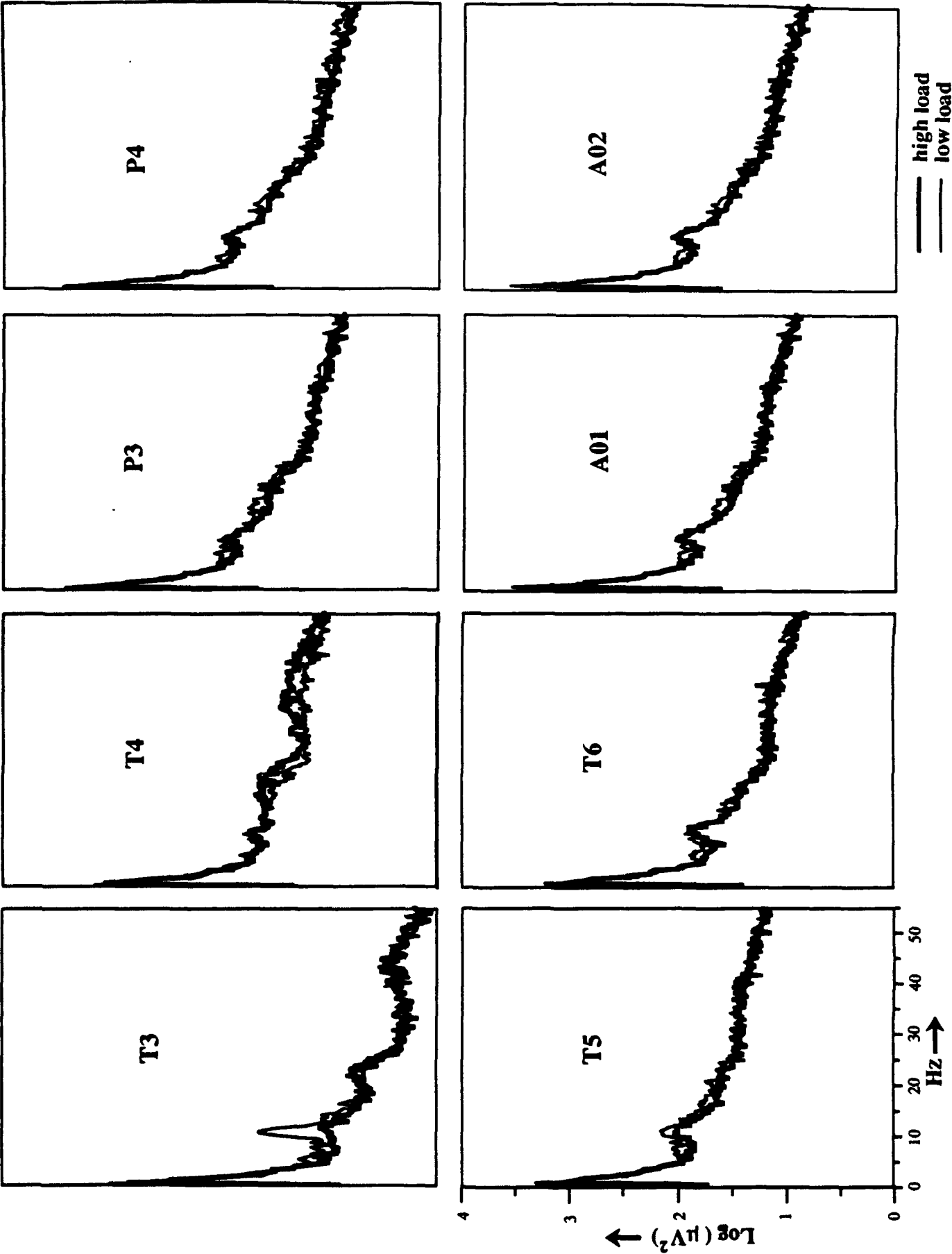
Figure 2b: Subject 2, average power spectra.
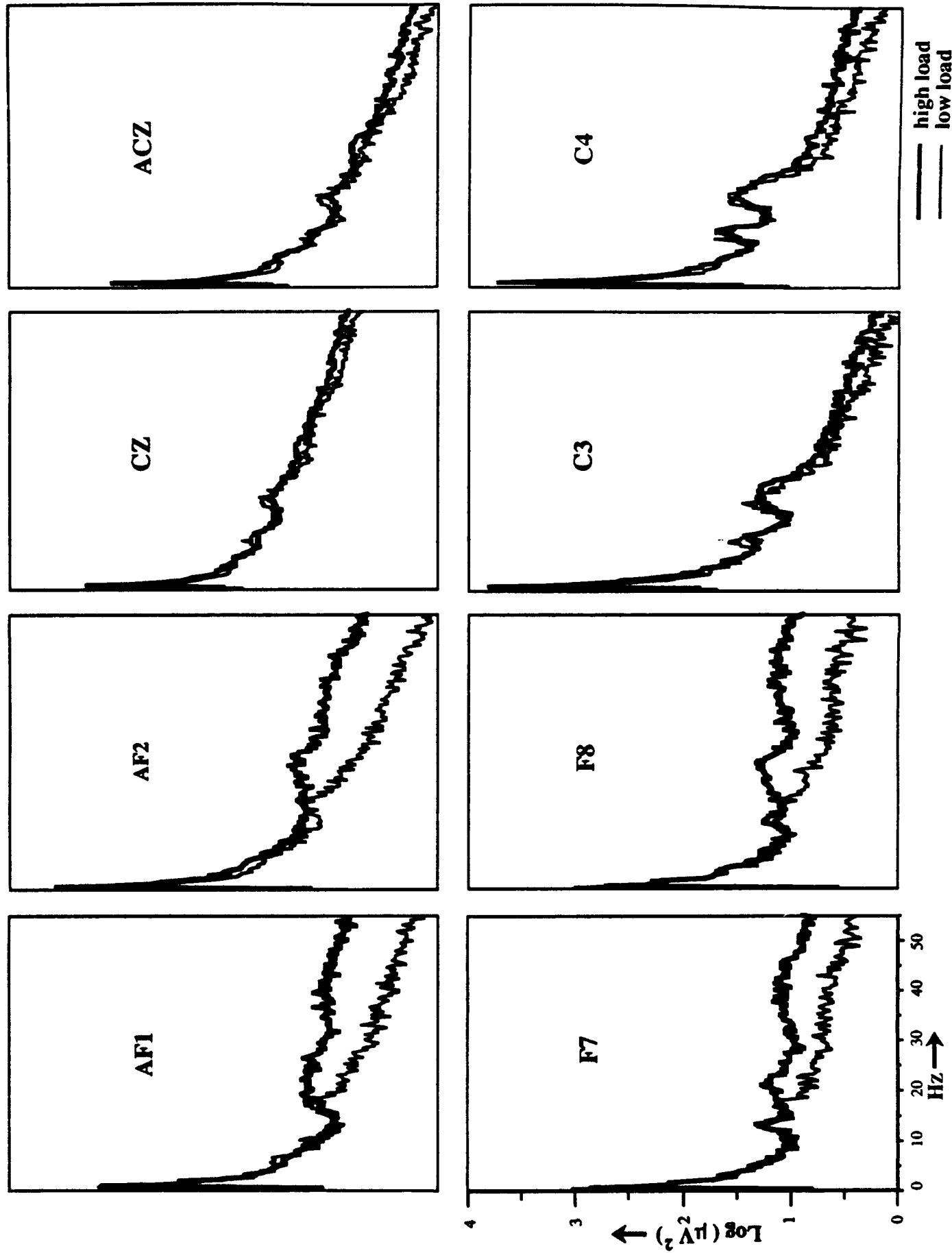
Figure 3a: Subject 3, average power spectra.
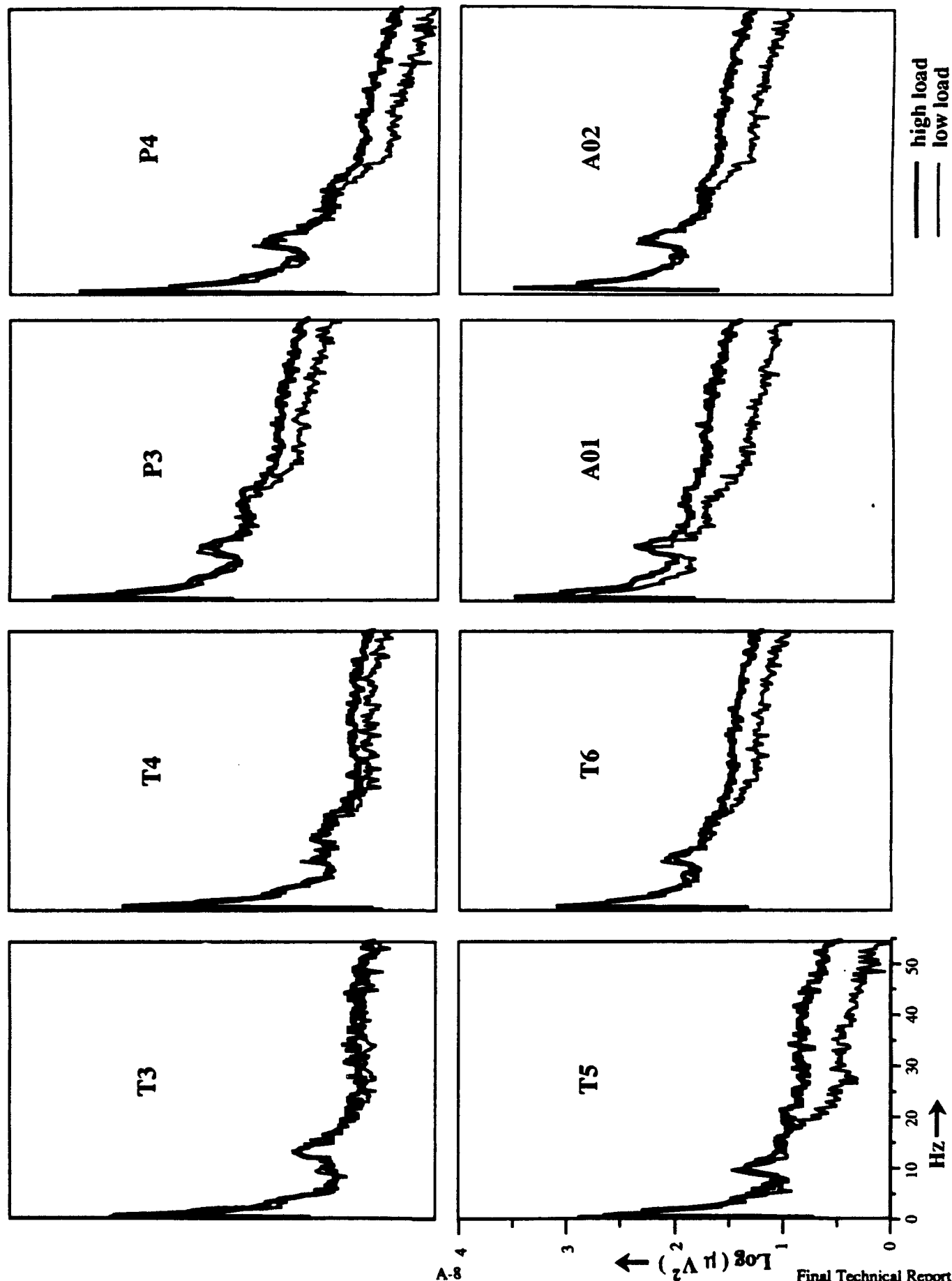
Figure 3b: Subject 3, average power spectra.
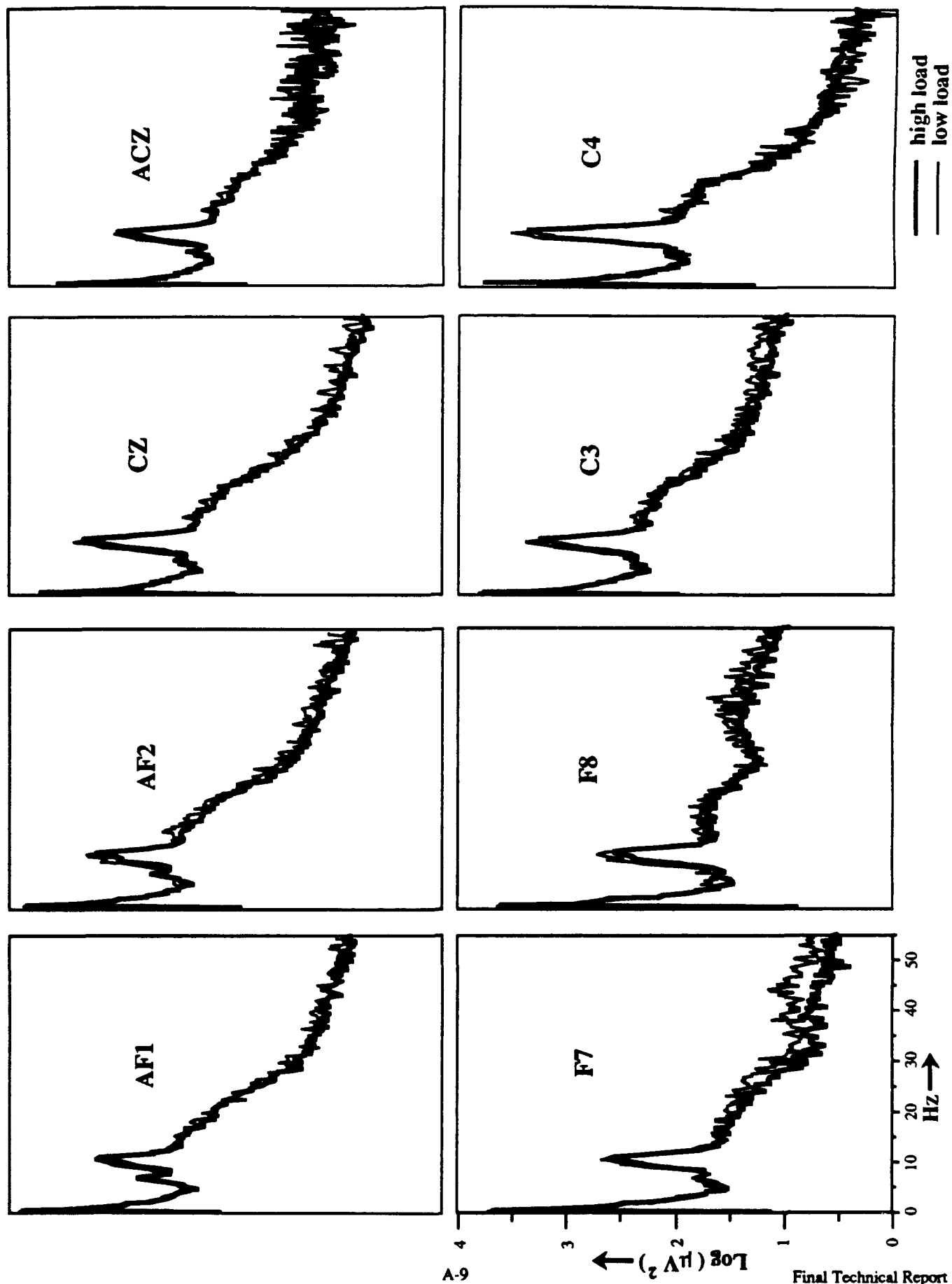
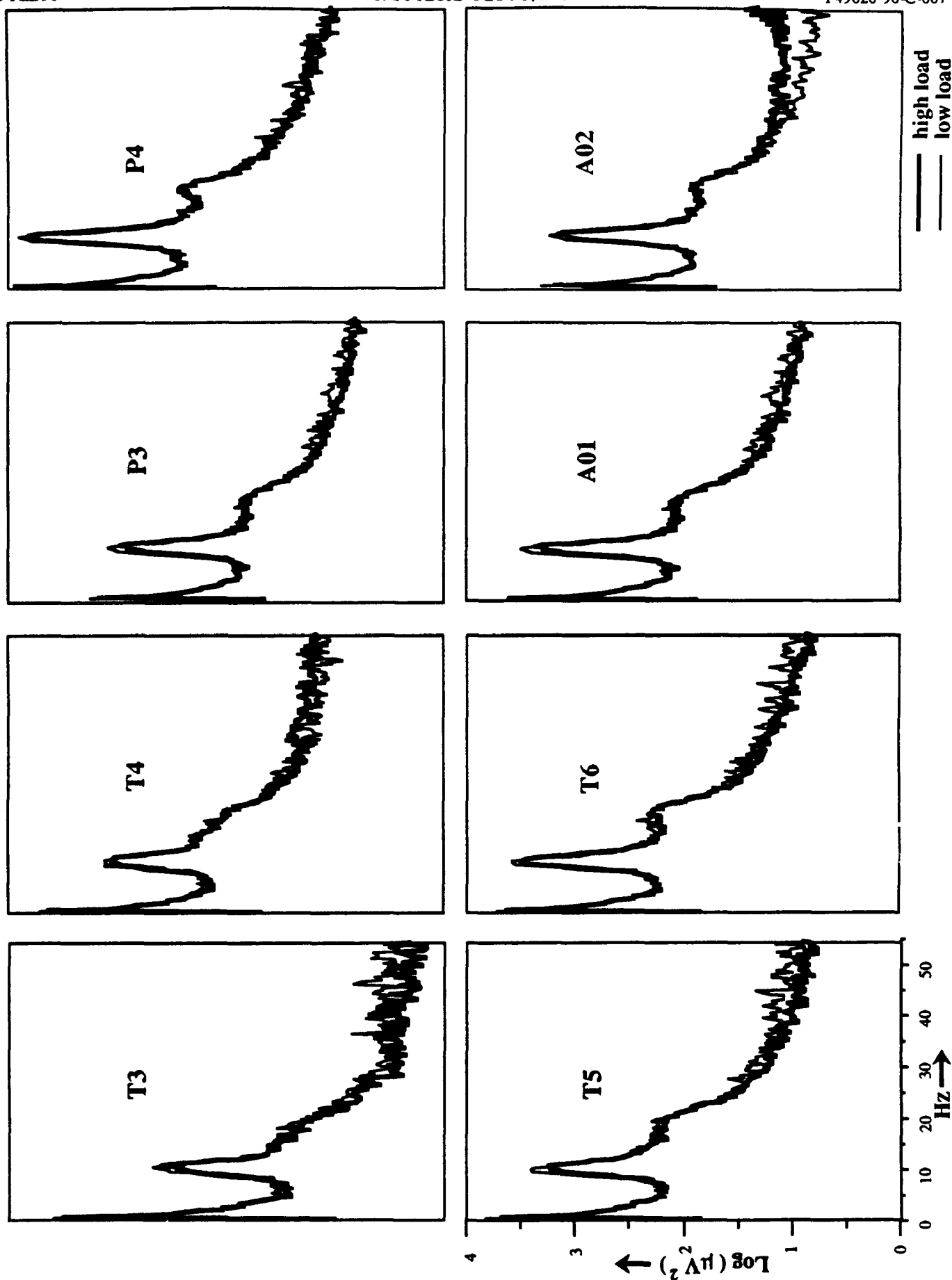Figure 4a: Subject 4, average power spectra.

Figure 4b: Subject 4, average power spectra.

high load
low load

$\leftarrow$ Log (μV^2)

Hz $\longrightarrow$

Figure 5: One-dimensional error probability of subject 3 for distinguishing zero-back from two-back conditions for the four frequency bands used. The single best discriminator was left occipital (AO1) EMG activity which correctly distinguished the workload conditions about 95% of the time (error probability 0.05). From this graph, we selected channels and frequency bands with error probability below 0.3 for the MM study and 0.35 for the CNS study for further investigation using neural networks. Some channels and frequency bands picked did not correspond to minima in error probability but were picked to have adequate representation across scalp locations and frequency bands.

Mental Workload Classification
Accuracy Using Heart Rate

Subject	HR mean	HR variance	HRV mean
1	58%	71%	61%
3	76%	62%	77%
4	69%	69%	72%

Table 1: Results of classifying high and low mental workload conditions using univariate heart rate measures. HRV = heart rate variance. (EKGs were not recorded for subject 2.)

Mental Workload Classification Accuracy
Using Clean Neurocognitive Signals

Subject	Test set classification accuracy	Most important features	Relative feature weights
1	97%	T5-theta	1
		T3-alpha	1
		AO2-theta	.54
2	100%	T3-alpha	1
		AF1-alpha	.73
3	100%	AO1-theta	1
4	92%	C3-beta	1
		T4-theta	.86
		AO2-alpha-var	.53
		ACZ-beta	.45
		T4-alpha	.38
		P3-beta	.37
		F8-alpha	.32
		T4-theta-var	.29

Table 2: Neural network classification results with sets of EEG features which are minimally sensitive to non-cognitive processes. Variances are labeled "var". For subject 4, principal components were used as inputs to the neural networks.

Mental Workload Classification Accuracy
Using Mixed Measure Signals

Subject	Test set classification accuracy	Most important features	Relative feature weights
1	95%	AO1-emg	1
		AO2-beta	1
		P3-emg	.54
2	100%	T3-alpha	1
		F7-alpha	.31
		F8-emg	.29
3	99%	F8-emg	1
		AO1-emg	.80
		AF1-beta	.42
4	94%	AO2-emg	1
		AF2-beta	.67
		T5-emg	.53

Table 3: Neural network classification results using both EEG and scalp EMG features.

APPENDIX B, Review of Existing Methods for Indexing Workload

Table of Contents

1. Subjective Estimates of Workload

This is the most common form of workload estimation. The measure is surprisingly consistent, but accuracy has been questioned (Gopher and Donchin, 1986; Kantowitz and Casper, 1988; Metcalfe, 1988). In many instances, subjective estimates are heavily influenced by the subject's conception of the task rather than how they actually responded to the task (Gopher and Braune, 1984). It is clear that these drawbacks are inherent to the measure considering that behavior is determined by both conscious and unconscious processes: workload is a function of both processes and, by definition, subjects can only report conscious experiences.

2. Overall Behavioral Performance

This measure is insensitive to gradations in mental workload primarily because it does not measure the use of resources until close to the capacity limits of at least one resource dimension. Performance can be held constant as workload changes dramatically (Kahneman, 1973). These deficits render the measure inappropriate for use in intelligent aircraft management systems since by the time performance decrements are observable, it may be too late.

3. Performance on Secondary Tasks

Many secondary tasks have been found to be sensitive to workload for many primary tasks (Kantowitz and Casper, 1988). Unfortunately, difficulties abound, many of them pointed out in a recent review (Gopher and Donchin, 1986; Boff and Lincoln, 1988). The relationship of the secondary to the primary task is crucial to the sufficiency of the former as a metric of workload. A major difficulty is to determine the relationship between the mental resources required by primary and secondary tasks and mental capacity. For example, a primary and secondary task may place high demands on motor output. This means the secondary task would be primarily sensitive to motor workload and insensitive to other mental resources such as memory. Nonlinear information processing capabilities of the brain increase the difficulty of quantifying the relationship: emergent resource demands may result from juxtaposing several tasks. Subjects may develop new, wholistic strategies to perform multiple tasks. Unless we have a model of cognitive function that completely specifies all sources of mental capacity and their interactions, we can only rely on secondary tasks to index mental workload in the context of a particular set of primary tasks.

Secondary task workload measures are inappropriate to use in intelligent aircraft management systems. Secondary task indices are very specific to their associated primary task(s), secondary tasks necessarily change the structure of primary tasks, and, by definition, secondary tasks introduce an additional informational load. These features would be highly unacceptable in a system in which operator performance is paramount.

4. Pupil Response

Pupil size has been observed to be a sensitive indicator of mental workload. Specifically, changes in pupil size have been correlated with changes in the load on memory storage and retrieval (Kahneman and Beatty, 1966; Beatty, 1982; Kahneman and Write, 1971; Stanners, 1972). Pupil size has also been observed to be correlated with stimulus probability (Qiyuan et al., 1985); this suggests that the measure might be useful in a manner similar to the P300 component of the evoked potential. Unfortunately, the sensitivity of pupil response was tested in circumstances where ambient lighting could be carefully controlled. The applicability of the method under conditions of variable ambient lighting awaits development of technical means to correct for ambient lighting.

5. Speech Signals

Parameters of speech signals including amplitude and pitch have been found to be sensitive to both high levels of stress (Kuroda et al., 1976) and subtle levels of workload (Brenner and Shipp, 1988). Other measures of speech quality, such as speech rate, energy distribution, amplitude shimmer, and frequency shimmer weakly distinguished workload conditions. By contrast, in a subcritical tracking task, heart rate was found to be a more reliable measure of tracking difficulty (Brenner and Shipp, 1988). Data reported in a preliminary study of speech qualities and workload (Alpert and Schneider, 1988) did not demonstrate a clear relationship. A major drawback is that speech samples may not be available at critical moments.

6. P300 and other brain evoked potentials (EPs)

Among brain evoked potential components, the P300 has received the most attention as a candidate metric for mental workload. This is a late evoked potential peak occurring about 300-600 msec following the reception of a stimulus which is preferably task relevant and relatively infrequent. Changes in the amplitude, latency or power integrated over a small interval of P300 have been observed to be sensitive to changes in workload. This sensitivity was observed for both active and passive P300 tasks where, in the former, subjects respond to the oddball events, and in the latter, no response was required (Kramer et al., 1981). The passive P300 findings suggest that a P300 metric could be used in systems requiring real-time monitoring of operator workload. Unfortunately, there are several problems. One problem is that, to effectively evoke P300s, oddball events must be task relevant stimuli (Duncan-Johnson and Donchin, 1977; Israel et al., 1980; Polich, 1989). For some primary tasks, this would be restrictive. Another problem is that P300s may only be sensitive to a restricted class of mental resources (Gevins and Cutillo, 1986). A series of studies have implicated P300 to be related to resources involved in evaluating stimuli (Kutas et al., 1977; Donchin et al., 1978; McCarthy and Donchin, 1981). These studies have demonstrated that latency of the P300 is independent of response selection and execution time. Indeed, it has been observed that when evaluation of stimuli was not an important variable in changing workload, P300 was insensitive to the change (Wickens et al., 1977). Thus, the specificity of the P300 may severely limit its utility. Finally, a P300 metric may not be robust over time. In early sessions of a flight simulation task with auditory stimuli for evoking P300, significant differences were observed in latencies and integrated power of P300 between two workload levels. This difference was substantially weaker in later sessions (Natani and Gomer, 1981).

Other EPs that have been examined include steady state EPs (Regan, 1980; Wilson and O'Donnell, 1986; Kramer et al., 1988), visual EPs evoked by task specific stimuli (Horst et al., 1984), and non-task specific visual stimuli (Trejo et al., 1987). Despite some promising initial results, several properties of EP metrics may limit their applicability in intelligent aircraft management systems. Single trial EPs have low signal-to-noise ratios and this may compromise reliability. Depending on the number of trials averaged, enhancing EPs by averaging may compromise time resolution. To evoke the necessary signals, either extraneous stimuli must be imposed or events that are naturally part of the operator environment must be used. The latter option would be optimal, but a workload measurement may not be available when most needed. Using simple extraneous stimuli would only test loading on primary sensory resources and complex extraneous stimuli would impose an additional informational load.

7. Continuously Measurable Physiological Signals

Physiological measures that are particularly well adapted for use in real-time intelligent aircraft management systems include background EEG, scalp muscle activity, EKG, and measures of eye blink and movement, skin conductance, and respiration. Signals are continuously available for all of these measures except eye movements and blinks and respiration. All measures do not require imposing additional workload and do not impose any structural requirements on the operator's tasks. A few

recent positive results using these measures are mentioned below.

7.1. Electrocardiograms (EKGs)

Heart rate (HR) has been observed to increase with rising workload over two related arithmetic tasks (Sharit and Salvendy, 1982) and over two levels of difficulty in a subcritical tracking task (Brenner and Shipp, 1988). Significant differences in heart rate variability and power spectra of IBIs have been observed over four workload levels of a dual stimulus memory task (Aunon et al., 1987). Heart rate variability power around 0.1 Hz was observed to decrease with increasing workload in a study using a 2 and 4 item memory task (Aasman et al., 1987). Respiratory sinus arrythmia (RSA) may be an indicator of parasympathetic activity relevant to workload (Gawron et al., 1989). RSA can be estimated from frequency analysis of EKG interbeat intervals (Porges, 1986). But Grossman (1983) showed the use of combined EKG and respiratory data to be superior in a context involving change in mental workload. T-wave amplitude has been related to cognitive or anticipatory stress (Heslegrave and Furedy, 1979).

7.2. Ongoing Electroencephalograms (EEGs)

EEG alpha and theta activity diminished with higher workload in a flight simulator task (Natani and Gomer, 1981). These results were robust over time compared to P300 workload related differences which diminished with practice. EEG high beta activity significantly increased as tasks changed from signal detection to signal recognition to memory to mental arithmetic (Kakizaki, Preprint). Peak alpha activity was observed to shift toward higher frequencies with higher workload for arithmetic and visual imagery tasks (Osaka, 1984). Theta suppression has been correlated with improved performance in vigilance tasks (Beatty et al., 1974; O'Hanlon and Beatty, 1977; Beatty and O'Hanlon, 1980). In-flight EEG data, primarily banded spectral power, indicate relevance to pilot cerebrocortical arousal in conditions of G-induced loss of consciousness (Lewis, et al., 1987) and fatigue (Howitt, et al., 1978). Reduction in alpha activity over one hemisphere with respect to the other is sensitive to secondary task performance performed in-flight during normal flight duties (Sterman, 1989).

7.3. Skin Conductance

Skin conductance has been observed to increase with increasing levels of semantic processing where phonetic processing is baseline (Cohen and O'Donnell, 1988). Lindholm and Cheatham (1983) found HR and skin conductance response (SCR) to be reliable indicators of short-term workload increases indexed by simulated aircraft carrier landings. In subsequent work, they found HR to be more stable than SCR in highly realistic simulated landing tasks (Lindholm, et al., 1984).

7.4. Eye Movement and Blink Measures

Eye blink rate has been observed to decrease with increases in visual processing demands (Stern and Skelly, 1984). Morris (1984ab, 1985) found blink rate, amplitude and duration were predictors of greater performance variability in straight and level and in maneuvering flight with fatigued pilots. This study, along with Skelly, et al. (1987) and Wilson, el al. (1987), suggests that eye blink waveform features such as mean duration may index the general state of cerebrocortical arousal of the pilot (Gawron, et al., 1989; see also Stern, et al., 1984; Fogarty and Stern, 1989).

7.5. Respiration Measures

Respiration rate (RR) by itself has shown only minor promise for indexing workload (Wierwille and Conner, 1983; Casali and Wierwille, 1984). Respiration data may be important for getting good measures of respiratory sinus arrythmia (RSA), an important measure for improving estimates of HR variability measures (Grossman and Wientjes, 1986).

7.6. Caveat

These physiological measures are usually individually tested for sensitivity to workload (e.g., Lindholm & Sisson, 1985; WierWille, et al., 1985; Casali and Wierwille, 1984; Lindholm, et al, 1984; Wierwille and Conner, 1983). The best performers vary from one experiment to the next, even within the same laboratory. Often, one or two are reported as doing well while two or three others do poorly. This has lead some to conclude that one or more of these measures are unreliable measures of workload (Johnson, 1980). In addition to variations in the type and quality of experimental design and execution (Gevins and Schaffer, 1980), lack of a precise definition of workload has also been cited as causing inconsistent results (Williges, et al., 1979; Aunon, et al., 1987).

8. References

Aasman J., Mulder G., Mulder L.J.M., (1987), "Operator effort and measurement of heart-rate variability," Human Factors, v29, pp161-170.

Alpert, M. and Schneider, S.J., (1988), "Voice-stress measure of mental workload," in Mental State Estimation 1987, NASA Conf. Publication 2504, pp155-160.

Aunon J.I., Kantowitz B.H., McGillem C.D., Plonski M.P., (1987), "Electromagnetic metrics of mental workload", AFOSR-85-0313 final report, Purdue University.

Beatty J., Greenberg A., Deibler W.P., and O'Hanlon J.F., (1974), "Operant control of occipital theta rhythm affects performance on a radar monitoring task", Science, v183, pp871-873.

Beatty J. and O'Hanlon J.F., (1980), "Operant control of posterior theta rhythms and vigilance performance: repeated treatments and transfer of training", In Birbaumer N. and Kimmel K.D. (Eds.), Biofeedback and Self_Regulation, Hillsdale, NJ, Erlbaum Press.

Beatty J., (1982), "Task-evoked pupillary responses, processing load, and the structure of processing resources", Psychological Bulletin, v91, pp276-292.

Boff, K.R., Lincoln, J.E., (Eds.), (1988), *Engineering Data Compendium: Human Perception and Performance, Vol. II*, John Wiley and Sons, New York.

Brenner M. and Shipp T., (1988), "Voice stress analysis", in Mental State Estimation 1987, NASA Conf. Publication 2504, pp363-376.

Casali John G. and Walter Wierwille (1984) "On the measurement of pilot perceptual workload: a comparison of assessment techniques addressing sensitivity and intrusion issues", Ergonomics, Vol 27, pp. 1033-1050.

Cohen R.A. and O'Donnell B.F., (1988), " Attention, effort, and fatigue: neuropsychological perspectives", in Mental State Estimation 1987, NASA Conf. Publication 2504, pp237-268.

Donchin E., Ritter W., and McCallum C., (1978), "Cognitive psychophysiology: the endogenous components of the ERP", in E. Callaway, P. Tueting, and S. Koslow (Eds.), Brain Event-Related Potentials in Man, New York, Academic Press.

Duncan-Johnson C.C. and Donchin E., (1977), "On quantifying surprise: the variation in event-related potentials with subjective probability", Psychophysiology, v14, pp456-467.

Fogarty, C. and Stern, J.A., (1989) "Eye movements and blinks: Their relationship to higher cognitive processes",

Gawron, V.J., Schiflett, S.G., and Miller, J.C., (1989) "Measures of inflight workload", in R. S Jensen (ed.), Aviation Psychology, Gower Technical Publishing (Aldershot, U.K.), Chap. 14.

Gevins, A.S. and Cutillo, B.A., (1986), "Signals of cognition, " in Lopes da Silva, F.H., (Ed.), Handbook of EEG Clin. Neurophys., (Revised Series Vol. 2), New York, Elsevier, pp335-381.

Gevins A.S. and Schaffer R.E., (1980) "A critical review of electroencephalographic (EEG) correlates of higher cortical functions", CRC Critical Reviews in Bioeng., v4, pp113-164.

Gopher D. and Braune R., (1984), "On the psychophysics of workload: why bother with subjective measures?", Human Factors, v26, n5, pp519-532.

Gopher D., Donchin E., (1986) "Workload--an examination of the concept" in Boff K.R., Kaufman L., and Thomas J.P. (Eds.), Handbook of Perception and Human Performance VII: Cognitive Processes and Performance, Chap. 41.

Grossman, P. and K. Wientjes (1986) "Respiratory sinus arrythmia and parasympathetic cardiac control: some basic issues concerning quantification, application and implications", in P. Grossman and K.H.L. Janssen (Eds.), Cardiorespiratory and Cardiosomatic Psychology (NATO ASI Series, Plenum Press), pp. 117-138.

Grossman, P. (1983) "Respiration, stress and cardiovascular function", Psychophysiology, Vol 20, pp. 284-300.

Heslegrave R.J. and J.J. Furedy (1979) "Sensitivities of HR and T-wave amplitude for detecting cognitive and anticipatory stress", Physiology and Behavior, 22, pp. 17-23.

Horst R.L., Munson R.C., Ruchkin D.S., (1984), "Event-related potential indices of workload in a single task paradigm," Proc. Human Factors Soc. -- 28th Ann. Mtg., San Antonio, TX, pp727-731.

Howitt, J.S., Hay, A.E., Shergold, G.R., and Ferres, H.M., (1978), "Workload and fatigue -- inflight EEG changes," Aviation, Space, and Env. Med., v49, n10, pp1197-1202.

Israel J.B., Chesney G.L., Wickens C.D., and Donchin E., (1980), "P300 and tracking difficulty: evidence for multiple resources in dual-task performance", Psychophysiology, v17, pp259-273.

Johnson L.C., (1980), "Use of physiological measures to monitor operator state," in Gomer F.E. (Ed.), Biocybernetic Applications for Military Systems, MDC E2191, McDonnell Douglas Astronautics Co , St. Louis.

Kahneman D., (1973), Attention and Effort, New York, Prentice Hall.

Kahneman D. and Beatty J., (1966), "Pupil diameter and load on memory", Science, v154, pp1583-1585.

Kahneman D. and Wright P., (1971), "Changes of pupil size and rehearsal strategies in a short-term memory task", Quarterly J. Exper. Psych., v23, pp187-196.

Kakizaki T., (Preprint), "Assessment of mental workload by means of occipital midline beta-2 amplitude".

Kantowitz, B.H. and Casper, P.A., (1988), "Human Workload in Aviation," in Wiener, E.L., Nagel, D.C., (Eds.), Human Factors in Aviation, pp263-303.

Kramer A., Wickens C.D., Vanasse L., Heffley E.F., and Donchin E., (1981), "Primary and secondary task analysis of step tracking: an event-related potentials approach", in R.C. Sugarman (Ed.), Proc. Twenty-Fifth Ann. Mtg. Human Factors Soc., Rochester, N.Y..

Kramer A.F., E. Donchin and D.D. Wickens (1988) "Event-related brain potentials as indices of mental workload and attentional allocation", NATO/AGARD Conference Proceedings, 432, pp. 14-1 to 14-15.

Kuroda I., Fujiwara O., Okamura N., and Utsuki N., (1976), "Method for determining pilot stress through analysis of voice communication", Aviation, Space, and Environmental Medicine, v47, pp528-533.

Kutas M., McCarthy G., and Donchin E., (1977), "Augmenting mental chronommetry: the P300 as a measure of stimulus evaluation time", Science, v197, pp792-795.

Lewis, N.L., McGovern, J.B., Miller, J.C., Eddy, D.R., and Forster, E.M., (1987), "EEG indices of G-induced loss of consciousness," in *Proc. NATO Advisory Group for Aerospace Res. and Devel. (AGARD), Electric and Magnetic Activity of the Central Nervous System: Research and Clinical Applications in Aerospace Medicine.*

Lindholm, E. and Sisson, N., (1985), "Physiological assessment of pilot workload in simulated and actual flight environments," *Behav. Res. Methods, Instr., and Comp.*, v17, n2, pp191-194.

Lindholm, E. and Cheatham, C.M., (1983), "Autonomic activity and workload during learning of a simulated aircraft carrier landing task," *Aviation, Space, and Env. Med.*, v54, n5, pp435-439.

Lindholm Ernest, Cary Cheatham, John Koriath and Thomas M. Longridge (1984) "Physiological assessment of Aircraft pilot workload in simulated landing and simulated hostile threat environments", (AFHRL-TR-83-49) Operations Training Division, Air Force Human Resources Laboratory.

McCarthy G. and Donchin E., (1981), "A metric for thought: a comparison of P300 latency and reaction time", Science, v211, pp77-80.

Metcalfe J., (1988), "Premonitions of insight predict impending error", J. Exper. Psych.: Learning, Memory and Cognition, v12, pp623-634.

Morris T.L. (1984a) " Electrooculographic indices of fatigue-induced decrements in flying related performance", (Technical Report). Brooks Air Force Base, TX: School of Aerospace Medicine.

Morris T.L. (1984b) " Electrooculographic measurement, fatigue, and variability of performance in simulated aircraft flight, Dissertation, College Station, TX: Texas A and M University", (Technical Report). Brooks Air Force Base, TX: School of Aerospace Medicine.

Morris T.L. (1985) "Electrooculagraphic indexes of changes in simulated flying performance", Behavior Research Methods, Instruments, and Computers, Vol 17, pp. 176-182.

Natani K., Gomer F.E., (1981), "Electrocortical activity and operator workload: a comparison of changes in the electroencephalogram and in event-related potentials", McDonnell Douglas Astronautics Company, St. Louis Division, Tec. Rpt. MDC-e2427.

O'Hanlon J.F. and Beatty J., (1977) "Concurrence of electroencephalographic and performance changes during a simulated radar watch and some implications for the arousal theory of vigilance", in Mackie R. (Ed.), Vigilance: Theory, Operational Performance, and Physiological Correlates, New York, Plenum Press.

Osaka M., (1984), "Peak alpha frequency of EEG during a mental task: task difficulty and hemispheric differences", Psychophysiology, v21, n1, pp101-105.

Polich J., (1989), "P300 from a passive auditory paradigm", Electroenceph. Clin. Neurophys., v74, pp312-320.

Porges, S.W. (1986) "Respiratory sinus arrythmia: physiological basis, quantitative methods and clinical implications" in P. Grossman and K.H.L. Janssen (Eds.), Cardiorespiratory and Cardiosomatic Psychology (NATO ASI Series, Plenum Press), pp. 101-115.

Qiyuan J., Richer F., Waggoner B.L., Beatty J., (1985), "The pupil and stimulus probability", Psychophysiology, v22, pp530-534.

Regan D., (1980), "Control system and physiological monitoring applications of steady-state evoked potentials," in Gomer F.E. (Ed.), Biocybernetic Applications for Military Systems, MDC E2191, McDonnell Douglas Astronautics Co., St. Louis.

Sharit J. and Salvendy G., (1982), "External and internal attentional environments: II. Reconsideration of the relationship between sinus arrythmia and information load", Ergonomics, v25, n2, pp121-132.

Skelly J.J., B. Purvis and G. Wilson (1987) "Fighter pilot performance during airborne and simulator missions: Physiological comparisons", Proceedings of the Advisory Group on Aerospace Research and Development (AGARD) Aerospace Medical Panel Symposium, Electric and Magnetic Activity of the Central Nervous System: Research and Clinical Application in Aerospace Medicine.

Stanners R.F., (1972), "The pupillary response to sentences: influences of listening set and deep structure", J. Verbal Learning and Verbal Behav., v11, pp257-263.

Sterman, M.B., (1989), "Psychophysiological studies of crew workload and alertness, final report," Northrop Corp. Project # 88-D-8077.

Stern J.A. and Skelly J.J., (1984), "The eye blink and workload consideration," Proc. Human Factors Soc., v28, pp942-944.

Stern John A., Larry C. Walrath and Robert Goldstein (1984) "The endogenous eyeblink", Psychophysiology, Vol 21, p. 22-23.

Trejo L.J., Lewis G.W., Blankenship M.H., (1987), "Brain activity during tactical decision-making: II. probe-evoked potentials and workload", NPRDC TN 88-12, Navy Personnel Research and Development Center.

Wickens C.D., Israel J., and Donchin E., (1977), " The event-related cortical potential as an index of task workload in adaptive man-machine systems", Proc. Mtg. of the Human Factors Soc., San Francisco, Oct.

Wierwille, W.W. and Connor, S.A., (1983) "Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator", Human Factors, Vol 25, pp. 1-16.

Wierwille Walter W., Mansour Rahimi and John G. Casali (1985) "Evaluation of 16 measures of mental workload res using a simulated flight task emphasizing mediational activity", Human Factors, Vol 27, pp. 489-502.

Willeges Robert C. and Walter W. Wierwille (1979) "Behavioral measures of aircrew mental workload", Human Factors, Vol 21, pp. 549-574.

Wilson, G.F., Purvis, B., Skelly, B., . .kamp, P., and Davis, I., (1987), "Physiological data used to measure pilot workload in actua! night and simulator conditions," In Proc. 31st Ann. Mtg. Human Factors Soc.

Wilson G.F. and O'Donnell .D., (1986), "Steady state evoked responses: correlations with human cognition," Psychophysiology, v23, pp57-61.